yang_seonhyeHW26

Seonhye Yang

4/2/2019

```
library(readr)
golf<- read.csv("golf_driver (1).csv")
```

##Question 1.a

```
attach(golf, warn.conflicts = F)

#mean for golfer
tapply(totdist, golfer, mean)
```

```
##      1      2      3      4      5      6
## 255.75 251.75 290.75 266.50 257.50 231.75
```

```
#mean for drivers
tapply(totdist, club, mean)
```

```
##        1        2        3        4
## 258.6667 262.3333 261.3333 253.6667
```

##Question 1.b

```
fit <- lm(totdist~factor(golfer)+factor(club))
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: totdist
##                Df Sum Sq Mean Sq F value    Pr(>F)
## factor(golfer)  5 7489.0 1497.80 79.0164 3.103e-10 ***
## factor(club)    3  270.7   90.22  4.7597   0.01589 *
## Residuals      15  284.3   18.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We get a F-value of 4.7597 and p-value of 0.01589 for drivers which is less than $\alpha = 0.05$ so we can say that it is significant. There is evidence that drivers can predict total distance.

Question 1.c

```
tukey <-TukeyHSD(aov(fit))
tukey
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = fit)
##
## $`factor(golfer)`
##       diff         lwr        upr     p adj
```

```
## 2-1  -4.00 -14.0022733   6.002273 0.7810800
## 3-1  35.00  24.9977267  45.002273 0.0000001
## 4-1  10.75   0.7477267  20.752273 0.0317690
## 5-1   1.75  -8.2522733  11.752273 0.9917109
## 6-1 -24.00 -34.0022733 -13.997727 0.0000147
## 3-2  39.00  28.9977267  49.002273 0.0000000
## 4-2  14.75   4.7477267  24.752273 0.0026637
## 5-2   5.75  -4.2522733  15.752273 0.4562039
## 6-2 -20.00 -30.0022733  -9.997727 0.0001221
## 4-3 -24.25 -34.2522733 -14.247727 0.0000130
## 5-3 -33.25 -43.2522733 -23.247727 0.0000002
## 6-3 -59.00 -69.0022733 -48.997727 0.0000000
## 5-4  -9.00 -19.0022733   1.002273 0.0902095
## 6-4 -34.75 -44.7522733 -24.747727 0.0000001
## 6-5 -25.75 -35.7522733 -15.747727 0.0000062
##
## $`factor(club)`
##          diff        lwr        upr     p adj
## 2-1  3.666667  -3.578093 10.9114268 0.4847685
## 3-1  2.666667  -4.578093  9.9114268 0.7174712
## 4-1 -5.000000 -12.244760  2.2447601 0.2352662
## 3-2 -1.000000  -8.244760  6.2447601 0.9778880
## 4-2 -8.666667 -15.911427 -1.4219066 0.0168483
## 4-3 -7.666667 -14.911427 -0.4219066 0.0363857
```

```
which(tukey$`factor(golfer)`[,4] <= 0.05)
```

```
## 3-1 4-1 6-1 3-2 4-2 6-2 4-3 5-3 6-3 6-4 6-5
##   2   3   5   6   7   9  10  11  12  14  15
```

```
which(tukey$`factor(club)`[,4] <= 0.05)
```

```
## 4-2 4-3
##   5   6
```

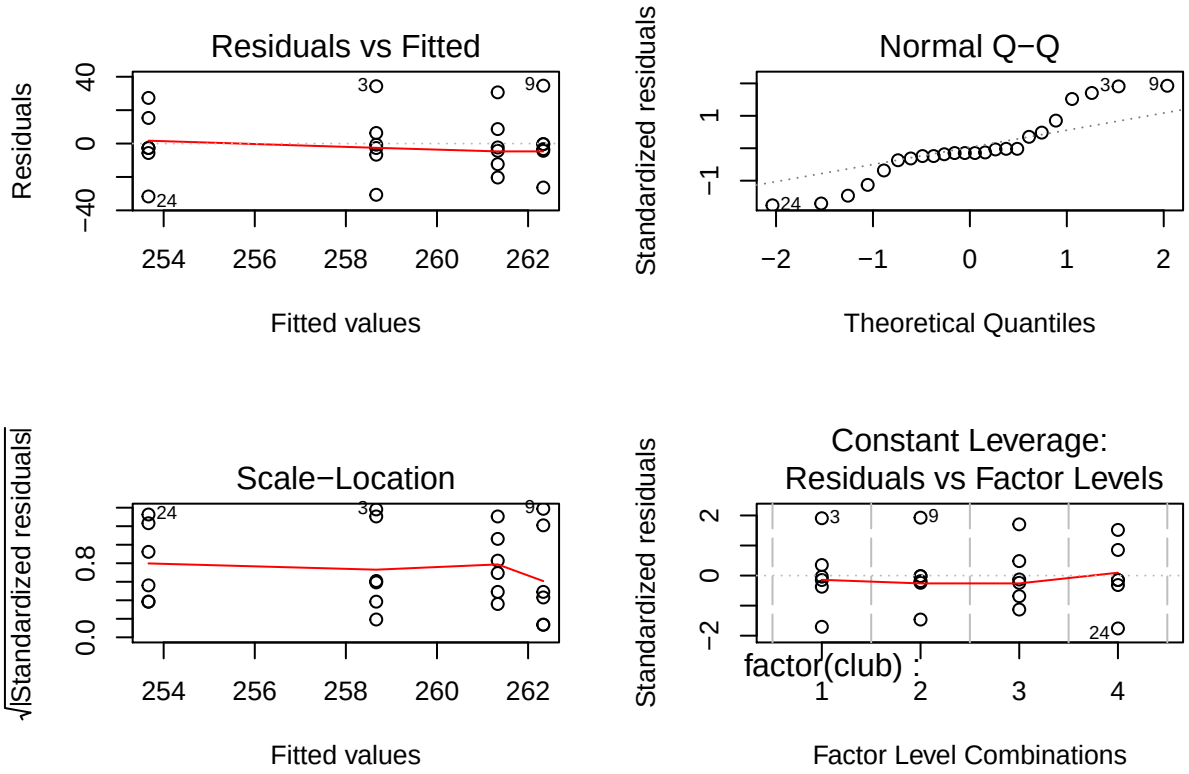
For golfers, 3-1 4-1 6-1 3-2 4-2 6-2 4-3 5-3 6-3 6-4 6-5 are significant because p-values are smaller than 0.05 and intervals don't include 0. And for drivers, 4-2 4-3 are significant because p-values are smaller than 0.05 and intervals don't include 0.

##Question 1.d

```
one_fit <- lm(totdist~factor(club))
anova(one_fit)
```

```
## Analysis of Variance Table
##
## Response: totdist
##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(club)  3  270.7   90.22  0.2321 0.8729
## Residuals    20 7773.3  388.67
```

```
par(mfrow=c(2,2))
plot(one_fit)
```



When no blocking occurs on golfers, drivers have a F-value of 0.2321 and a p-value of 0.8729. The p-value is greater than 0.05 which means drivers is not significant for predicting total distance.

Looking at Residuals vs Fitted, there seems to be a linear relationship. Looking at Normal Q-Q, the residuals don't seems to be normally distributed so the constant variance is violated and Scale-Location shows that residuals are spread equally along the ranges of predictors which doesn't violated normality. Lastly, there doesn't seem to be any outliers or potential influential points.

It's just constant variance that is violated.

##Question 1.e

```
TukeyHSD(aov(one_fit))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = one_fit)
##
## $`factor(club)`
##          diff       lwr      upr     p adj
## 2-1  3.666667 -28.19152 35.52485 0.9880903
## 3-1  2.666667 -29.19152 34.52485 0.9953221
## 4-1 -5.000000 -36.85818 26.85818 0.9709176
## 3-2 -1.000000 -32.85818 30.85818 0.9997483
## 4-2 -8.666667 -40.52485 23.19152 0.8707114
## 4-3 -7.666667 -39.52485 24.19152 0.9058489
```

The results show that none of the driver pairs are significant. All of the p-values are greater than 0.05 and all of the confidence intervals have 0.